晋江文学城原创文学网站用户画像研究

■ 宋雪雁 张梦笛
吉林大学管理学院 长春 130022

摘 要：[目的/意义]探究晋江文学城原创文学网站用户属性，构建晋江文学城原创文学网站用户画像，在拓展用户画像理论应用场景的同时，帮助网站更细致地了解用户文化消费行为，从而优化网站运营，提升用户粘性，为其他原创文学网站的用户研究提供参考。[方法/过程]以晋江文学城原创文学网站用户为研究对象，运用因子分析法、K-Means聚类分析和Word Art词云工具，构建晋江文学城原创文学网站用户画像，分析用户兴趣及行为规律。[结果/结论]确立晋江文学城原创文学网站用户画像标签体系，构建4类用户画像，并以不同类型用户画像为参照，提出晋江文学城原创文学网站优化策略。

关键词：用户画像 原创文学网站 晋江文学城 用户行为
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.23.007

1 引言

随着我国互联网普及率的提高，文学、音乐、视频等娱乐内容生态持续向好，网络文学用户数量再创新高。国内原创文学网站作为文化消费的主要阵地，其用户群体十分庞大，访问量非常可观，对经济发展的影响力也在不断扩大[1]。作为最具代表性和经济价值的原创文学网站，晋江文学城网站的注册用户已超4 000万，访问者来自全球近200个国家和地区，日均页面浏览量在4亿以上，主体业务为实体出版、电子版权采购及海外版权合作，拥有简体言情出版市场70%以上的小说版权，同时不断推进影视、动漫、广播剧等领域的项目合作，打造了众多知名IP，有效推动了网络文学的传播[2]。用户画像最早由国外学者A. Cooper提出并应用于软件行业，后逐渐渗透到其他行业[3]。近年来，用户画像研究作为国内图书馆学、情报学等领域的研究热点，文献发表量持续增长，研究内容不断丰富，众多学者及其团队已取得代表性研究成果。但目前针对原创文学网站用户画像的研究还未引起足够重视，因此，本研究以晋江文学城原创文学网站用户为对象，结合因子分析法、K-Means聚类分析和Word Art词云工具，识别晋江文学城网站用户特征，确立晋江文学城网站用户画像标签体系，构建晋江文学城原创文学网站用户画像，并以此为参照提出网站优化策略。本文相较于前人研究，一方面有助于丰富研究主题，拓展用户画像理论的应用场景，弥补原创文学网站用户研究的不足；另一方面通过深入分析网站用户的阅读和消费特征，可以优化网站运营，提升用户粘性，更好地促进用户的文化消费，实现网站的需求分析与价值创造，也为其他原创文学网站的用户研究提供参考。

2 文献回顾

2.1 原创文学网站相关研究

国内外数据库中相关文献的检索结果显示，国外鲜有研究涉及原创文学网站，国内对于原创文学网站的研究主要集中在出版、运营、传播、版权等领域，其中出版发行方面包括图书出版[4]、网络文学出版[5]、付费阅读[6]等内容，运营方面以竞争力评价[7]、SWOT分析[8]、盈利模式[9]、全版权运营[10]、IP运营[11]、微博营销[12]等为主题开展研究，传播方面涉及群体传播[13]、网络文学传播[14]、网络编辑[15]等内容，版权保护以信息网络传播权[16]和数字版权保护[17]等研究为主。目前很少有将原创文学网站用户作为研究对象的系统研究。作为原创文学网站的代表，晋江文学城相关研究主要涉及文学出版[4-5]、女性主义[18]、战略分析[8]、营销手段[11-12]等内容，缺少对用户进行有针对性的研究。

作者简介：宋雪雁（ORCID：0000-0003-1955-6685），教授，博士生导师，E-mail：songxueyan@126.com；张梦笛（ORCID：0000-0001-5447-8155），硕士研究生。
收稿日期：2020-06-12 修回日期：2020-09-07 本文起止页码：63-74 本文责任编辑：易飞

2.2 用户画像概念

在英文文献中，通常以“User Persona”“User Profile”或“User Portrait”来表示用户画像的含义。“User Persona”由交互设计之父 A. Cooper 于 1998 年提出，是设计过程中最重要的匹配用户需求的虚拟角色，是以真实数据为基础构建的共享特定目标的用户模型[3]。“User Profile”的定义是一组描述用户属性的结构化数据集[19]，一般被理解为用户信息标签化，即从不同维度用户属性中获取有价值的用户信息，提取用户标签，构建用户画像[20-21]，主要研究学者为 B. Atote、金小月和韩梅花等。T. Tang 和 K. Huang 等认为用户画像“User Portrait”是通过提供用户历史数据来模拟用户特征，进而建立起用户需求或概念模型[22-23]。国外学者通常使用“User Persona”和“User Profile”来表示用户画像，二者的侧重点各有不同，“User Persona”更多地是推断、勾画用户目标并传递用户诉求，“User Profile”则是以建立描述用户信息的符集为重点。国内对用户画像的解释可分为两类，一类是定性分析用户需求，抽象用户画像，另一类是定量获取用户信息，构建用户标签[24]，但国内有些研究也存在概念模糊与理论深度欠缺等问题[25]。

2.3 用户画像构成要素

从用户画像的构成要素来看，国外学者强调了不同领域用户画像信息内容的拓展，T. Miaskiewicz、C. Lerouge、C. Moeckl 等认为用户画像应由描述、图片、技术对应物[26]等静态信息[27]和兴趣偏好、行为态度、情景事件等动态信息[28]构成。P. Kamthan、S. Ouaftouh 和 C. Schäfer 等将用户画像的构成要素分为生理、技能、目标等强制性信息和信仰[29]、情感[30]、组织、环境等可选择性信息[31]。国内学者则是在应用场景不断拓展的基础上，进一步补充用户的属性类型，并对用户标签体系划分的合理性进行验证。陈志明、刘海鸥、刘速等学者提出了自然属性、社交属性、兴趣属性、能力属性[32]、情景属性、会话属性[33]和资源属性[34]；刘应杰、李映坤和吴加琪对原始数据层、事实层、模型预测层、业务层[35-36]和特征偏好层[37]进行了研究。尽管用户画像的构成要素在不断完善，但应用场景仍存在空白，用户动态画像和用户实时画像等研究还未得到重点关注。

2.4 用户画像研究方法

关于用户画像研究方法，A. Cooper 等提出了利用虚拟情节描述用户行为特征的七步人物角色法[38]，L. Nielsen 提出以目标为导向并将叙事场景引入用户画像构建的十步人物角色法[39]，A. Madureira 等对不同研究阶段可应用的用户画像研究方法进行总结，包括调查访谈、上下文查询、焦点小组、德尔菲法和可用性测试等[40]。高广尚将用户画像研究方法按照基于设计与思维、基于本体或概念、基于主题或话题、基于用户兴趣或偏好、基于用户行为或日志和基于多维或融合的标准进行划分[24]。关梓骜和陈添源十分重视技术手段在用户画像中的应用，对主题模型、统计聚类、深度学习、规则定义等数据挖掘技术进行了充分研究[41-42]。总体而言，国外学者首先开展用户画像的相关研究，取得一系列研究成果的同时提出多种典型的定性、定量研究方法。国内用户画像研究处于高速发展阶段，各类技术手段应用层出不穷，但仍存在数据安全、案例支撑等问题，研究成果的可重复验证性与可持续发展面临考验[25]。

3 晋江文学城原创文学网站用户画像构建

用户画像是以用户为中心，合理划分用户维度，提取用户属性标签，利用用户真实数据结构化地表现用户属性特征的过程。构建完整的晋江文学城原创文学网站用户画像，需要充分了解网站特点及用户特征，建立符合研究需要的晋江文学城网站用户画像标签体系，借助因子分析与聚类分析获得具有不同属性特征的晋江文学城网站用户画像类型。

3.1 用户画像标签提取与数据获取

标签是对用户属性信息抽象后获得的语义化、短文本化的特征标识[43]。本研究通过对相关文献中不同领域用户画像构成要素的梳理，结合晋江文学城网站特点和网站用户行为分析，在众多不同的构成要素中选取符合研究对象特征的基础属性、兴趣属性和行为属性来划分晋江文学城网站的用户属性。受网站运营特点的影响，用户的主要行为特征表现为阅读和消费，因此行为属性由用户阅读属性和用户消费属性共同构成。基于用户属性划分对晋江文学城原创文学网站用户画像标签体系进行预设，共提取 51 个标签，见图 1。其中晋江文学城网站用户的基础属性由用户等级、性别、年龄、教育背景、职业和收入水平 6 个标签构成。用户的兴趣属性由网站偏好和文章偏好等 12 个标签组成，对网站的偏好被细分为选择晋江文学城网站的动机、网站内容建设、广告推送数量与时长；对文章的偏好被细分为首要关注的网站分区、文章的属性、进度、风格、时代、类型、章节及全文字数。晋江文学城网站用户的行为属性由 33 个标签组成，其中阅读属性

由用户的阅读时段、阅读场所、阅读设备、阅读渠道、阅读方式、阅读时长、使用时长、计划性阅读频率、反复性阅读频率、选择性阅读频率、评论打分频率、文章分享频率、论坛交流频率、评审任务频率、观看广告频率、阅读数量、收藏数量、评论数量等 21 个标签构成。用户消费属性由充值频率、订阅频率、包月频率、充值渠道、订阅渠道、支付渠道、充值金额、订阅金额、打赏金额、消费参考和消费产品等 12 个标签构成。

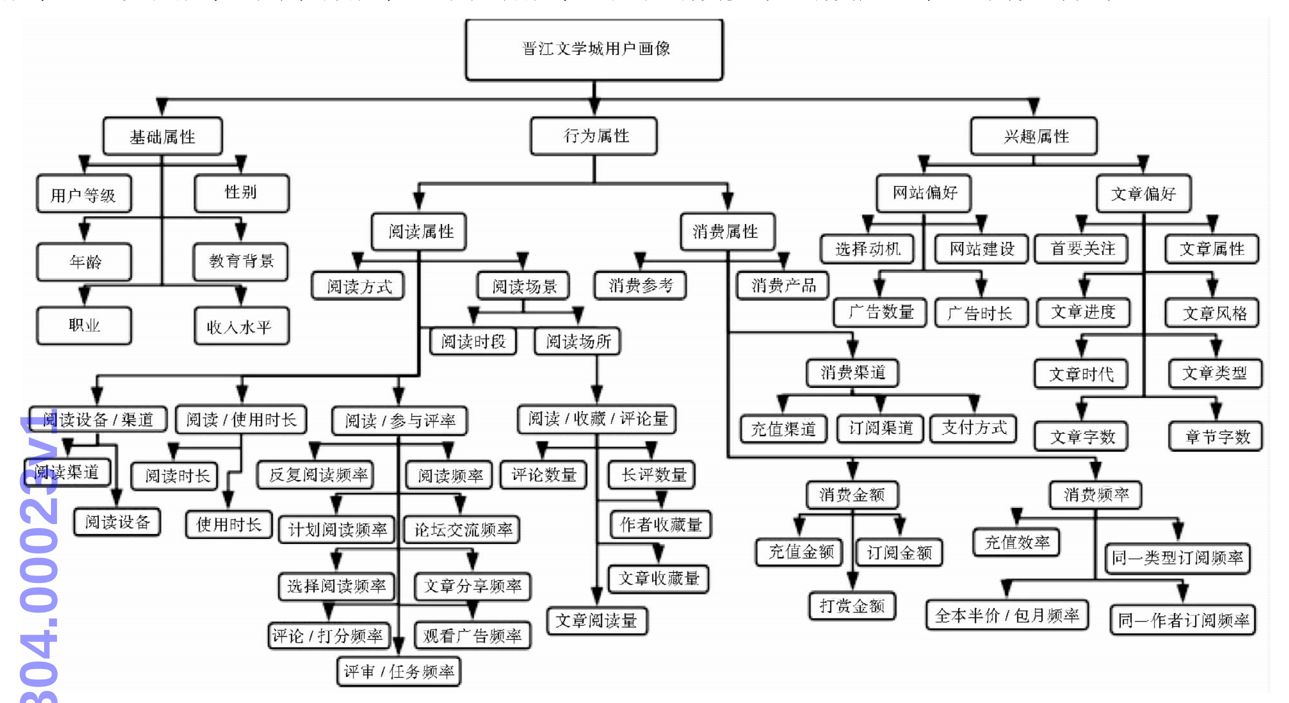


图 1　预设的晋江文学城原创文学网站用户画像标签体系

本研究以提取的晋江文学城原创文学网站用户画像标签为基础制作调查问卷，并根据用户对问卷的反馈作出修改，问卷共分为用户基础、兴趣、阅读和消费属性 4 个研究部分。面向晋江文学城网站 QQ 群、微信群和网站论坛中的用户发放在线调查问卷，自 2020 年 3 月 3 日至 2020 年 3 月 24 日共回收问卷 1 363 份，有效问卷量为 1 177，有效率达到 86.4%。本研究将采用描述属性值频数的方式来解释用户的人口统计变量、兴趣变量和不适合赋值的行为变量，以不违背原始数据采集意义为前提，对于适合赋值的用户行为变量借助李克特量表来表示用户的行为程度。

3.2　数据描述性分析与信度效度分析

3.2.1　描述性统计分析

（1）描述用户基础属性。如表 1 所示，此次调查中晋江文学城网站普通用户占比 26%，因为普通用户由未付费用户和消费金额未达到等级的付费用户组成，因此网站的付费用户数量不低于 74%，表明多数用户愿意接受并使用网站的付费阅读模式。女性用户与男性用户的性别比接近 4:1，说明晋江文学城网站更符合女性群体的审美偏好，也证明了网站“女性网络文学原创基地”定位的准确性。晋江文学城网站用户的年龄结构整体呈年轻化趋势，18－29 岁的用户接近样本的 9/10，青年人成为网站用户的主体。从用户的教育背景和职业分布来看，网站用户具有良好的教育背景，85% 为本科及以上学历，职业主要为学生、企事业单位工作者和自由职业者等。89% 的网站用户月收入在 5 000元以下，用户群体普遍收入较低。基于本次调查结果可知，晋江文学城网站的核心受众和消费主体为女性用户，因此网站需要将其作为目标受众，识别用户特征，迎合用户需求，同时不断扩大受众范围，发展高收入用户群体，提高经济效益以保证网站的长久运行。

（2）描述用户兴趣属性。由表 2 可知，45% 的用户受网站知名度、影视游戏、作者水平和作品质量的影响成为晋江文学城网站的用户，46% 的用户的选择动机为消磨时间和舒缓压力。关于网站建设，71% 的用户对网站的设计排版和版权保护表示关注，说明这两部分是网站调整和建设的关键，同时页面运行超载或瘫痪、操作程序受限以及客服声誉不佳等问题也亟待解决。广告是网站获得经济收入的重要来源，76% 的用户并不排斥广告推送，当以虚拟货币为物质奖励时，61% 的用户愿意接受 6 秒以上的广告时长，因此晋江文学城网站可以选择面向部分用户进行适时适量的广

表 1 晋江文学城网站用户基础属性数据

用户等级	人数(个)	性别	人数(个)	年龄(岁)	人数(个)	教育背景	人数(个)	职业	人数(个)	月收入(元)	人数(个)
普通用户	304	女	965	1 – 17	57	初中及以下	29	学生	903	500 及以下	289
消费用户	493	男	212	18 – 29	1047	高中	68	党政机关工作者	12	501 – 2 000	464
初级 vip 用户	283			30 – 39	41	中专/中职/中技	19	事业单位工作者	78	2 001 – 5 000	291
高级 vip 用户	97			40 – 49	16	专科(高职/高专/高技)	65	企业工作者	94	5 001 – 8 000	92
				50 – 59	14	本科	869	个体经营者	17	8 001 – 10 000	18
				60 及以上	2	硕士研究生	119	自由职业者	54	10 000 以上	23
						博士研究生	8	外出务工人员	8		
								农林牧渔劳动者	1		
								退休人员	4		
								失业/无业人员	6		

告推送。在选择作品时,47% 的用户以各类排行榜为依据,说明在网站现有的内容分区中,作品排行榜对用户来说具有重要的参考价值,网站需要保证排行榜的公平性与合理性,不断满足用户需要,方便用户选择。用户通过网站自有的分类系统检索阅读作品,关于文章属性,91% 的用户喜欢原创性文学作品,这一点十分符合网站原创文学的定位,同时证明了晋江文学城原创文学网站存在的意义与价值;在文章进度方面,73% 的用户喜欢已完结的作品,说明多数用户在主观或客观因素的影响下对连载作品的兴趣较小;就文章风格而言,轻松或正剧的文风更受读者欢迎;从时代背景来看,喜欢近现代文章背景的用户占比为 43%,说明用户更喜欢能够引起共鸣或产生代入感的作品。关于文章类型,用户对爱情类作品,悬疑科幻等“烧脑”类作品,漫画风格显著、受众年轻化的轻小说体裁以及充满中国传统元素的仙侠武侠文化作品表现出明显的倾向性。晋江文学城网站用户对章节字数的选择较为分散,39% 的用户选择每章 3 000 – 5 000 字,在文章长度方面,64% 的用户对 50 万字以内的作品感兴趣,长篇作品考验了用户的耐心。

表 2 晋江文学城网站用户兴趣属性数据

网站建设	人数(个)	广告数量(条)	人数(个)	广告时长(秒)	人数(个)	首要关注	人数(个)	章节字数(字)	人数(个)
设计排版	569	从不推送	281	0 – 5	454	作者专区	179	3 000 以内	196
操作难度	102	1 – 3	660	6 – 15	574	作品库	164	3 000 – 5 000	459
运行速度	147	4 – 6	163	16 – 30	129	版权专区	42	5 000 – 8 000	238
服务态度	95	7 – 9	34	31 – 60	14	文章标题	39	8 000 – 10 000	106
版权保护	264	10 及以上	39	60 以上	6	文章类型	195	10 000 以上	178
						文章封面	4		
						各类排行榜	554		

选择动机	人数(个)	文章偏好	人数(个)	文章类型	人数(个)	文章字数(万字)	人数(个)
消磨业余时间	357	原创性	1 069	爱情	791	30 以内	399
舒缓压力	180	衍生性	108	武侠	250	30 – 50	360
研究需要	21	连载中	316	奇幻	292	50 – 80	223
网站知名度高	117	已完成	861	仙侠	357	80 – 100	110
受影视、游戏影响	37	近代现代	507	游戏	132	100 以上	85
喜欢的作者在网站更文	197	架空历史	288	传奇	152		
使用时间长、个人情怀	68	古色古香	244	科幻	262		
文章质量高、章节注水少	183	幻想未来	138	童话	106		
寻求他人认同、寻找共同话题	17	轻松	631	惊悚	169		
		悲剧	60	悬疑	418		
		爆笑	157	剧情	504		
		正剧	281	轻小说	309		
		暗黑	48	古典衍生	172		
				东方衍生	103		
				西方衍生	59		

(3)描述用户行为属性。如表 3 所示，晋江文学城网站用户对阅读时段和阅读场所的选择在一定程度上反映了用户的阅读习惯，网站用户的主要阅读时间为 18:00 - 24:00，习惯性阅读地点为居住场所。90% 的用户以手机为常用访问设备，并且访问手机版 WAP 和 APP 的用户占比也达到 90%，说明随着网络技术与设备性能的提高，手机在晋江文学城网站发展与阅读推广中的影响力日渐扩大。网站用户对阅读方式的选择各有不同，71% 的用户喜欢逻辑性和连贯性阅读，29% 的用户喜欢跳跃式阅读或选择性阅读。在充值渠道方面，受苹果手机持有人数、苹果应用商店手续费等因素影响，半数的用户以安卓 APP 为充值渠道，使用苹果 APP 充值的用户仅占 15%。由于手机用户数量的增长和 APP 订阅可享受的高级 VIP 用户折扣，61% 的用户以手机 APP 端为主要订阅渠道。受用户使用习惯、网站便捷程度和手续费价格等因素的影响，78% 的用户以支付宝或微信为主要支付方式。另外针对部分用户拒绝消费的现状，晋江文学城网站有必要采取一定措施来吸引用户消费，增加网站收益。在消费参考方面，60% 用户将文案和文章基本信息作为参考重点，文案内容的精彩性和文章基本信息的准确性需要得到有效保证。版权输出与版权合作为晋江文学城网站创造了巨大的经济价值，78% 的用户存在购买网站授权产品的行为，这种行为有利于影视、出版、广播等领域的共同发展。

表 3　晋江文学城网站用户行为属性数据

阅读时段	人数(个)	阅读场所	人数(个)	阅读设备	人数(个)	阅读渠道	人数(个)	阅读方式	人数(个)
0 - 6 点	58	居住场所	676	手机	1065	手机版 WAP	381	从头到尾按章阅读	839
6 - 9 点	54	学习/办公场所	136	台式电脑	28	手机 APP	674	跳跃式阅读	129
9 - 12 点	102	交通工具	49	笔记本电脑	41	网页版 PC	115	阅读核心章节	74
12 - 14 点	81	任意场所	316	平板电脑	43	其他	7	阅读感兴趣的章节	128
14 - 18 点	85							阅读结尾	7
18 - 24 点	797								
充值渠道	人数(个)	订阅渠道	人数(个)	支付方式	人数(个)	消费参考	人数(个)	消费产品	人数(个)
不充值	267	不订阅	250	不支付	234	文案	424	无	264
WAP 端	104	WAP 端	137	支付宝	498	文章基本信息	282	出版书籍	572
安卓 APP	584	APP 端	720	微信扫码	342	文章评论	179	影视	376
苹果 APP	171	PC 端	70	支付宝花呗	77	VIP 前的免费章节	221	网络剧	332
PC 端	51			银行卡	13	部分 VIP 章节	71	游戏	123
				神州充值卡	7			动漫	255
				固话/电信手机账户	3			漫画	214
				移动/联通手机账户	3			广播剧	174
								有声读物	123
								周边	95

3.2.2　信度效度分析

适合赋值的晋江文学城网站用户行为变量包括阅读、使用时长、计划性阅读频率、反复性阅读频率、选择性阅读频率、评论打分频率、文章分享频率、论坛交流频率、评审任务频率、观看广告频率、阅读数量、收藏数量、评论数量、充值频率、订阅频率、包月频率、充值金额、订阅金额、打赏金额等 23 个变量，利用 SPSS 软件检验这些变量的信度和效度并作出评价。如表 4 和表 5 所示，Cronbach's Alpha 系数是 0.891，信度较高；KMO 值是 0.902，Bartlett 球形检验的 P 值在 0.001 以下，达到显著要求，说明可以选用因子分析法进行研究。

表 4　可靠性统计

Cronbach's Alpha	基于标准化项的 Cronbach's Alpha	项数
.891	.896	23

表 5　KMO 和 Bartlett 球形检验

KMO 取样适切性量数		.902
Bartlett 球形检验	近似卡方	11 035.744
	自由度	253
	显著性	000

3.3　用户行为变量因子分析

本研究通过因子分析对晋江文学城网站用户的行为变量进行检验与分析，抽取适合数据聚类的公因子，并对预设的晋江文学城原创文学网站用户画像标签体

系进行修正。首先利用主成分分析处理网站用户众多的行为变量,选择"最大方差法"和"旋转解",将特征值大于1、因子系数大于0.5作为获得公因子的判断标准。如表6所示,共有5个公因子特征值大于1,其方差累积贡献率为58.76%,因此可以抽取5个公因子来代替23个用户行为变量。

表 6　晋江文学城网站用户行为变量因子分析解释的总方差

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	7.253	31.534	31.534	7.253	31.534	31.534	4.164	18.104	18.104
2	2.291	9.962	41.496	2.291	9.962	41.496	3.315	13.632	31.736
3	1.635	7.109	48.605	1.635	7.109	48.605	2.680	11.650	43.386
4	1.319	5.737	54.341	1.319	5.737	54.341	1.937	8.421	51.807
5	1.016	4.419	58.760	1.016	4.419	58.760	1.599	6.953	58.760
6	.946	4.114	62.875						

通过观察晋江文学城网站用户行为变量旋转成分矩阵,可以了解各变量的因子系数以及各变量分别属于哪个公因子。如表7所示,阅读数量的因子系数小于0.5,说明用户每月的阅读数量对用户行为无显著影响,因此删除这一变量。其余22个用户行为变量分别属于5个不同的公因子,用户行为属性的原始分类标准不再适用于这些公因子和所含变量,因此需要修正预设的网站用户行为标签,重新配置用户属性,从而得到符合网站特点和用户特征的晋江文学城原创文学网站用户画像标签体系。

表 7　晋江文学城用户行为变量旋转成分矩阵

变量	成分				
	1	2	3	4	5
长评数量	.783				
论坛交流频率	.732				
文章分享频率	.727				
评论数量	.703				
评论/打分频率	.682				
作者收藏数量	.633				
文章收藏数量	.537				
阅读数量					
订阅金额		.840			
充值金额		.804			
打赏金额		.744			
充值频率		.618			
全本半价/包月频率		.615			
使用时长			.665		
阅读频率			.605		
阅读时长			.591		
同一类型订阅频率				.687	
同一作者订阅频率				.678	
广告频率				.649	
评审/任务频率				.505	
选择阅读频率					.768
反复阅读频率					.540
计划阅读频率					.536

作为晋江文学城网站用户行为特征的重要表现方式,每个公因子需要以所含变量为依据重新命名。

公因子1涉及论坛交流、文章分享、评论打分的频率以及收藏和评论的数量等7个变量。这些变量体现了网站用户的社交活跃度、作者作品关注度、情感表达程度和资源需求程度,所以公因子1被定义成互动需求因子。

公因子2涉及充值、订阅、打赏金额和充值、包月频率等5个变量。它们代表了网站用户固定的消费额度和消费规律,所以公因子2被定义成消费能力因子。

公因子3涉及阅读、使用时长和阅读频率3个变量。这些变量体现了用户在网站使用过程中访问频率等会话习惯[44]以及持续使用所投入的时间成本,所以公因子3被定义成会话成本因子。

公因子4涉及订阅、评审任务、观看广告的频率等4个变量。这些变量说明了网站用户的消费意识与文章类型、作者等因素间的相关性以及用户从网站奖励活动中获得的消费动力,所以公因子4被定义成消费驱动因子。

公因子5涉及计划性、反复性、选择性阅读频率3个变量。这些变量在旋转成分矩阵中仍属于同一公因子,说明三者内部关联紧密,并且作用于网站用户的阅读效率和阅读效果,所以公因子5被定义成阅读质量因子。

修正后的晋江文学城原创文学网站用户画像标签体系见图2。

运用信度效度分析检验修正后的用户行为变量,其中Cronbach's Alpha系数是0.888,KMO值是0.898,P值小于0.001,信度和效度较好,修正后的晋江文学城原创文学网站用户行为属性划分得到了有效验证。

宋雪雁,张梦笛.晋江文学城原创文学网站用户画像研究[J].图书情报工作,2020,64(23):63－74.

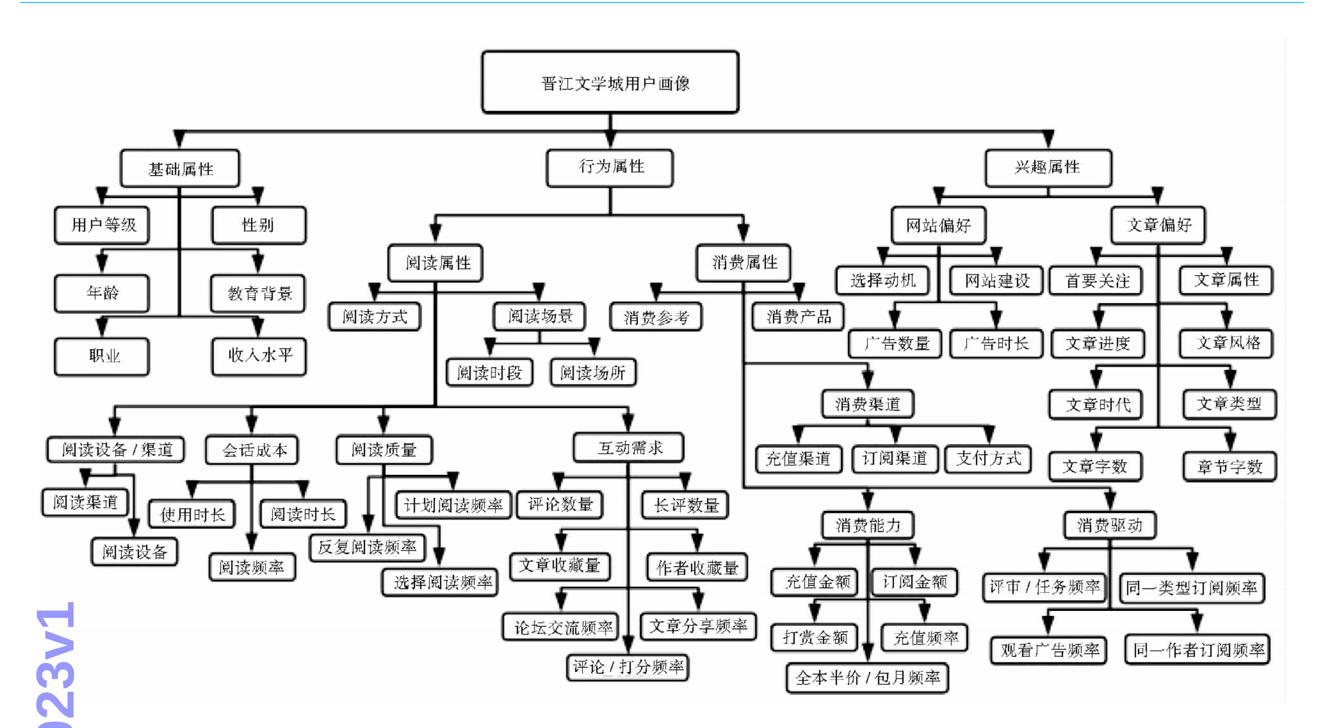


图2 修正后的晋江文学城原创文学网站用户画像标签体系

3.4 聚类分析

采用K-Means算法对通过主成分分析获得的5个公因子进行聚类,并通过判别式分析对聚类结果进行检验。人工设置三组聚类,聚类数为4－6,由表8可知,各因子对聚类均有贡献,组内数据之间存在显著性差异。指定聚类数6的结果中存在个案数极少的聚类,因此聚类数不宜设定为6。利用判别式分析检验聚类数4和5的有效性,可知两组sig值均小于0.001,聚类结果有效,聚类数5的正确分类概率略高于聚类数4,但差距较小,观察发现聚类数4的因子间差异更加显著,最终选择4作为最佳聚类数。

表8 聚类数为4－6的ANOVA表和判别式分析

公因子	聚类数4		聚类数5		聚类数6	
	F	Sig	F	Sig	F	Sig
互动需求	522.909	.000	353.382	.000	301.717	.000
消费能力	58.404	.000	14.401	.000	44.181	.000
会话成本	313.597	.000	250.566	.000	211.311	.000
消费驱动	128.008	.000	309.507	.000	217.754	.000
阅读质量	116.047	.000	215.235	.000	115.938	.000
威尔克 Lambda 函数检验		.000		.000		
正确分类概率	96.2%		96.8%			

在聚类数为4的情况下确定初始聚类中心,随着迭代次数的增加,聚类中心不断变化并逐渐实现收敛,最后获得最终聚类中心和不同聚类包含的个案数,即此次调查中不同用户类型的数量,见表9。

表9 最终聚类中心和每个聚类中的个案数目

公因子	聚类			
	1	2	3	4
互动需求	－.29167	－.39070	－.35186	1.67934
消费能力	－.13532	.54119	－.39058	.29062
会话成本	－.78115	1.08650	.10635	－.14977
消费驱动	－.60939	－.11985	.64176	.06331
阅读质量	－.37614	－.50760	.63787	.13603
个案数目	360	250	369	198

如表10所示,通过计算每个聚类组内的平均值获得晋江文学城原创文学网站最终的用户分类,并以平均值为依据分析不同类型网站用户的行为特征,根据用户行为特征命名网站用户画像类型。第一类用户各公因子均值较低,说明这类用户无明确资源需求、缺少强烈的消费需求和消费动力,网站使用率、社交活跃度

和消费额度较低，因此可称作体验型用户；第二类用户的消费能力和会话成本因子均值较高，说明这类用户的网站使用率和消费成本较高，已经形成固定的阅读和消费习惯，因此可称作成熟型用户；第三类用户的互动需求和消费能力因子均值较低，消费驱动和阅读质量因子均值较高，说明尽管这类用户的社交活跃度和消费能力处于较低水平，但有明确的资源需求且消费欲望强烈，提高消费的可能性较大，因此可称作提升型用户；第四类用户的互动需求、消费能力和阅读质量因子均值较高，说明这类用户重视阅读效果和情感表达，有明确资源需求，消费额度高且消费规律固定，因此可称作社交型用户。样本中 62% 的用户为体验、提升型用户，38% 的用户为社交、成熟型用户，这一情况证明了网站用户消费水平的可提升性以及制定网站优化策略的必要性。

表 10　平均值

个案聚类编号	互动需求	消费能力	会话成本	消费驱动	阅读质量
1	-.291 665 1	-.135 322 4	-.781 148 0	-.609 390 4	-.376 137 5
2	-.390 704 2	.541 191 6	1.086 501 2	-.119 853 4	-.507 598 0
3	-.351 855 4	-.390 582 6	.106 349 5	.641 759 0	.637 873 0
4	1.679 344 1	.290 621 9	-.149 772 8	.063 307 3	.136 029 7

4　晋江文学城原创文学网站用户画像类型

网站不同类型用户之间行为特征差异显著，基于聚类结果补充用户的基础属性、兴趣属性和不可赋值的行为属性，利用 Word Art 构建属性完整的晋江文学城原创文学网站用户画像标签云。可赋值用户行为标签大小取决于因子平均值，用户基础、兴趣和不可赋值的行为属性标签则由调查问卷中出现频数较高的问项来表示。最终得到 4 种晋江文学城网站的用户类型，分别是体验型用户、提升型用户、社交型用户和成熟型用户。

4.1　体验型用户画像

体验型用户画像如图 3 所示，这类用户多数是普通和消费等级，女性，18 – 29 岁，本科或研究生教育背景，学生、企事业单位工作者和自由职业者，月收入普遍低于 2 000 元；无明确的网站选择，易受设计排版等影响，以消耗时间和排解压力为目的访问网站，排斥广告推送；关注网站的各类排行榜，对每章 5 000 字以内、整篇低于 50 万字、原创性、已完结、轻松或正剧、近现代或架空历史的作品感兴趣，喜欢爱情、悬疑、剧情、仙侠等类型的文章；网站使用时长不超过一年，阅读频率较低，以手机 APP 和手机版 WAP 为主要访问渠道，时间一般为 18 – 24 点，时长少于 1 小时；阅读方式以按章节顺序阅读为主，有时会有选择性阅读，计划阅读和反复阅读的频率较低；很少参与论坛交流与作品分享，收藏和评论量低；对广告和评审等活动缺乏积极性，偶尔订阅同一作者或同一类型的作品，购买出版书籍和影视等授权产品，多数情况下拒绝充值或订阅。这类用户与网站的联系并不持久且极易流失，用户体验有待提高。

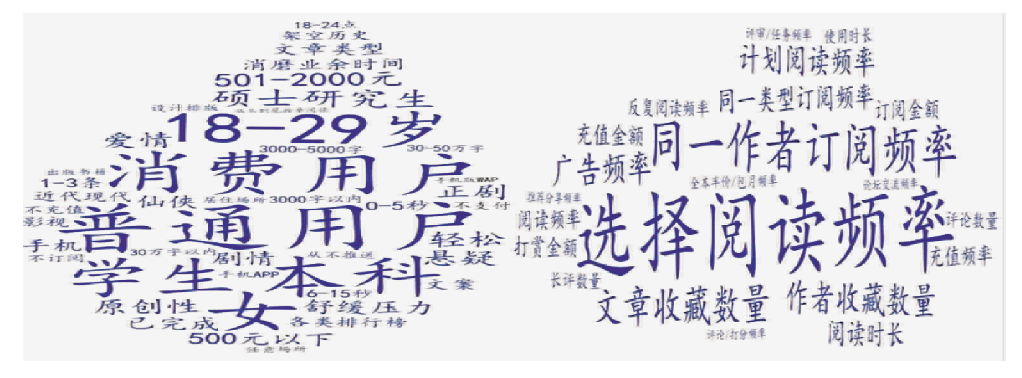


图 3　晋江文学城网站体验型用户画像标签云

4.2　提升型用户画像

提升型用户画像如图 4 所示，这类用户通常为普通或消费等级，女性，18 – 29 岁，高中或本科学历，职业是学生或企业工作者，月收入低于 2 000 元；受网站作者和作品吸引，重视设计排版和版权保护，对广告的接受程度为每天 3 条以内且不超过 15 秒；关注网站的排行榜，喜欢原创性、已完结、以近现代或架空历史为背景、轻松和正剧文风、每章 3 000 – 5 000 字、50 万字以内的作品，对爱情、悬疑、剧情、轻小说等类型文章感兴趣；网站使用时长一般不超过 3 年，阅读频率为每周多次、每次时长在 2 小时以内，时间通常在 18 – 24 点，主要访问渠道为 APP 或手机 WAP 端；时常出现有计

宋雪雁，张梦笛.晋江文学城原创文学网站用户画像研究[J].图书情报工作,2020,64(23):63－74.

划的、有选择的或反复的阅读行为;并不热衷于评论、收藏和分享等活动;充值金额少于 100 元,月订阅和月打赏金额均不超过 15 元,购买的授权产品一般为出版书籍、影视剧、网络剧等,消费频率较低;以文案和文章基本信息为消费参考,采用评审和广告等方式获得虚拟货币奖励,经常订阅同一作者或同一类型的作品。尽管提升型用户与体验型用户在用户活跃度和消费水平等方面存在相似的行为特征,但提升型用户有特定资源内容需求和明显的消费意图,发展空间存在可提升性。

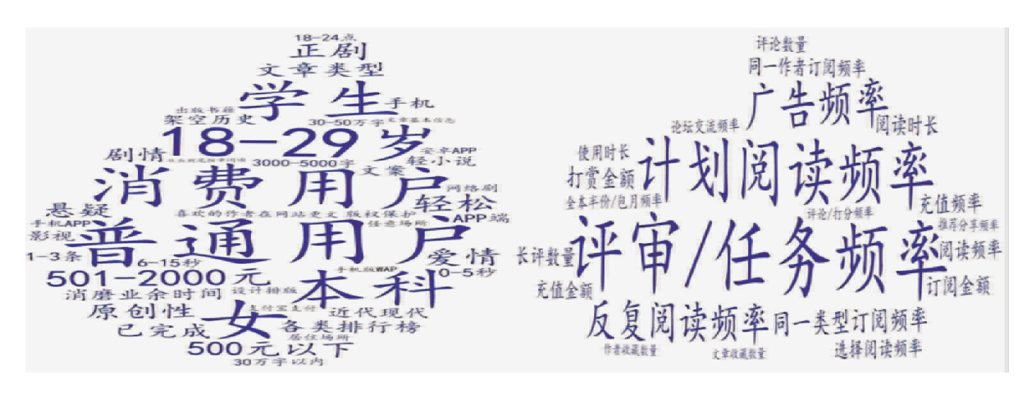


图 4　晋江文学城网站提升型用户画像标签云

4.3　社交型用户画像

社交型用户画像如图 5 所示,这类用户一般为消费用户或初级 VIP 用户,女性为主体,年龄在 18－29 岁之间,本科学历,职业为学生或企事业单位工作者,月收入为 501－5 000 元,高于其他三类用户;受影视、游戏、他人或话题的影响,通过网站实现缓解压力、消磨时间等目的,重视网站的设计排版,愿意接受广告推送;喜欢有知名度和话题性的,多种属性、进度、风格、时代和类型的作品,兴趣分布较为平均;使用时间在 6 年以内,阅读频率较高且阅读时间长,时间与地点的选择较为分散,存在选择性、计划性和反复性阅读行为;社交活跃度高,经常参与交流和分享,有强烈的表达欲和分享欲,收藏评论数量高,存在感明显;以微信或支付宝为支付手段,保持多月或每月一次的充值习惯,消费金额较高,月打赏金额为 6－30 元;积极参与评审和广告等兑换活动,订阅同一类型或作者的作品,经常购买各类网站授权产品。社交型用户与网站联系密切,愿意尝试各种内容和活动,用户体验丰富。



图 5　晋江文学城网站社交型用户画像标签云

4.4　成熟型用户画像

成熟型用户画像如图 6 所示,这类用户通常是消费和初级 vip 等级,女性,18－29 岁,本科或研究生,以学生为主要群体,月收入在 5 000 元以内;受作品质量、作者能力、个人情怀等影响,变更阅读网站的可能性较小,关心网站的设计排版和版权保护工作,对广告推送较为排斥;以文章类型和各类排行榜为参考,偏好原创性、已完结、风格轻松、近现代或古代背景、每章少于 8 000 字、总字数低于 50 万字的爱情、悬疑、仙侠、剧情类作品;使用时间多数在 10 年以内,阅读频率高,每周或每天多次阅读,阅读时间为 18－24 点,持续 1－4 小时;已形成阅读习惯,会制定阅读计划,偶尔表现出选择和反复阅读行为;很少使用网站的非阅读类功能,评审评论、交流分享等行为出现的频率较低,但文章和

作者的收藏量较高,专注于网站的内容;消费水平高于其他三类用户,订阅同一类型和同一作者的作品的频率较高,选择支付宝作为支付渠道,充值金额总计高于 100 元,订阅、打赏金额月均 6 - 30 元,消费产品一般为出版书籍、影视和广播剧等。成熟型用户对网站的内容和规则熟谙于心,用户体验丰富,具有较高的稳定性。



图 6　晋江文学城网站成熟型用户画像标签云

5　晋江文学城网站优化策略

针对晋江文学城网站不同类型的用户画像提出相应的网站优化策略,有利于网站提高服务水平与服务质量,增加用户粘性,促进用户的文化消费,实现网站的价值创造。

5.1　针对晋江文学城网站体验型用户:扩大宣传、丰富资源

通过分析体验型用户画像,提出网站优化策略:①整合媒体资源,多网站联动,扩大宣传和推广,提高网站知名度,避免竞争中“同质化”趋势,保持自有特色,做好系统维护工作,畅通访问渠道;②健全审核机制,遏制低俗文化传播,安装舆情监控系统,塑造良好的企业形象;③保证资源的数量与质量,合理规划网站分区,简化业务流程,提高运行速度;④开展免费试读、新人折扣、分享和邀请福利等活动,引起用户阅读兴趣,促进用户消费。

5.2　针对晋江文学城网站提升型用户:调动情绪、精准推送

通过分析提升型用户画像,提出网站优化策略:①通过社交平台和站内交流版块等途径密切用户联系,增强用户参与感,以物质或精神奖励的形式有效调动用户情绪,激励用户消费;②营造良好的写作氛围,规范作者签约管理制度,保证高质量内容产出,重视作者权益的维护;③根据用户的检索、阅读、收藏等历史信息,定位用户核心需求,提供符合用户偏好的精准推送;④设置相关作品推荐和阅读提醒功能,提高检索效率,方便用户阅读,提供个性化的阅读和消费服务,增加用户粘性。

5.3　针对晋江文学城网站社交型用户:粉丝经济、选择多样

通过分析社交型用户画像,提出网站优化策略:①开放站内交流通道,组织论坛、书展、见面会等线下活动[8],提供社交机会;②借鉴弹幕系统、添加段评和标记信息等互动方式,实现技术突破;③探索 IP 运营模式,授权影视游戏等改编,合理利用“粉丝效应”,正确引导粉丝群体为相关文化产品买单,为粉丝群体的二次创作提供生存和发展空间;④以用户需求为导向,鼓励作者创新,为用户提供多样化的信息资源和服务选择,从而提高用户忠诚度。

5.4　针对晋江文学城网站成熟型用户:深化服务、精品共享

通过分析成熟型用户画像,提出网站优化策略:①深化服务内涵,创新服务模式,建立合理的评价与监督机制,定制优质服务;②提高版权保护意识,严肃处理抄袭、盗版等问题,确保用户隐私和支付安全,承担社会责任;③开展业务能力培训,提升编辑服务性与专业性,细化文章分类,多角度评价作品质量,提供精品推荐[11];④完善用户等级制度,设置优先权和专属福利,提高用户忠诚度,避免稳定用户流失。

6　结语

构建晋江文学城原创文学网站用户画像并不是为了固化网站用户的兴趣与行为模式,而是通过增进对

网站用户的了解,更好地促进网站发展。本研究通过设计晋江文学城网站用户画像标签体系,构建了 4 类晋江文学城网站用户画像,分别是体验型用户、提升型用户、社交型用户和成熟型用户。最后,针对不同类型的用户提出了相应的网站优化策略,包括扩大宣传力度,畅通访问渠道,提供丰富内容;调动用户情绪,规范作者管理,实现个性化营销;线上线下联动,发挥“粉丝效应”,满足多样需求;深化服务内涵,重视版权保护,实现精品资源共享。晋江文学城原创文学网站用户画像研究不仅有利于网站运营方实现需求分析与价值创造,也为其他原创文学网站的用户画像研究提供参考。

参考文献:

[1] 中国互联网络信息中心. 第 44 次中国互联网络发展状况统计报告[R/OL]. [2020－02－01]. http://www.cac.gov.cn/2019-08/30/c_1124938750.htm.
[2] 晋江文学城. 关于晋江[EB/OL]. [2020－05－01]. http://www.jjwxc.net/aboutus/#fragment-31.
[3] COOPER A. The inmates are running the asylum: why high-tech products drive us crazy and how to restore the sanity[M]. 2nd Ed. New York: Sams Publishing, 2004.
[4] 张佩佩. 晋江文学城女性图书出版研究[D]. 西安:陕西师范大学,2013.
[5] 孙嘉咛. “耽美文学”出版研究[D]. 成都:西南交通大学,2013.
[6] 孙国钰. 论原创文学网站 VIP 付费阅读制度的生存环境[J]. 出版发行研究,2014(6):70－73.
[7] 高雪. 我国原创文学网站竞争力评价研究[D]. 长沙:湖南大学,2014.
[8] 吕融融. 原创文学网站多元化经营的 SWOT 分析[D]. 武汉:华中师范大学,2012.
[9] 余海燕. 网络原创文学网站的赢利模式分析——以起点中文网为例[J]. 出版发行研究,2015(2):46－48.
[10] 杨寅红. 盛大文学全版权运营模式研究[D]. 兰州:兰州大学,2013.
[11] 杜瑜. 晋江文学城 IP 运营研究[D]. 西安:陕西师范大学,2018.
[12] 项菡颖. 原创文学网站的微博营销研究[D]. 武汉:华中师范大学,2014.
[13] 边妮. 网络文学的群体传播研究[D]. 长春:吉林大学,2017.
[14] 谢易霖. 基于“5W”模式的网络文学传播研究[D]. 南宁:广西大学,2015.
[15] 张雅琴. 原创文学网站网络编辑的工作内容研究[D]. 保定:河北大学,2015.
[16] 陈建芳. 原创文学网站的信息网络传播权保护研究[D]. 武汉:华中科技大学,2017.
[17] 孙妍峰. 我国网络原创文学的版权问题研究[D]. 保定:河北大学,2014.
[18] 韦宝华. 后现代语境下网络耽美小说的“女性向”叙事研究[D]. 南宁:广西大学,2019.
[19] ATOTE B S, SAINI T S, BEDEKAR M, et al. Inferring emotional state of a user by user profiling[C]// Proceedings of the 2016 2nd international conference on contemporary computing and informatics. Noida: IEEE, 2016: 530－535.
[20] 金小月. 个性化档案数据服务——用户画像[J]. 卷宗,2018(15):35.
[21] 韩梅花,赵景秀. 基于“用户画像”的阅读疗法模式研究——以抑郁症为例[J]. 大学图书馆学报,2017,35(6):105－110.
[22] TANG T T, YIN Z Y, ZOU Y. A method for telecom user portrait modeling[C]//Research institute of management science and industrial engineering. Proceedings of 2017 5th international conference on frontiers of manufacturing science and measuring technology. Beijing: Atlantis Press, 2017: 1198－1203.
[23] HUANG K H, DENG Y S, CHUANG M C, et al. Static and dynamic user portraits[J]. Advances in human-computer interaction, 2012: 1－16.
[24] 高广尚. 用户画像构建方法研究综述[J]. 数据分析与知识发现,2019,3(3):25－35.
[25] 单轸,邵波. 国内图书馆领域用户画像研究的冷思考:困境与出路[J]. 图书馆学研究,2019(5):8－13,30.
[26] MOECKL C. From user-centred design to security: building attacker personas for digital banking[C]//Proceedings of the 10th Nordic conference on human-computer interaction. 2018: 01－03.
[27] MIASKIEWICZ T, KENNETH A K. Personas and user-centered design: how can personas benefit product design processes? [J]. Design studies, 2011, 32(5): 417－430.
[28] LEROUGE C, MA J. User profiles and personas in consumer health technologies[C]// Proceedings of the annual Hawaii international conference on system sciences. Honolulu: IEEE, 2010: 2433－2442.
[29] OUAFTOUH S, ZELLOU A, IDRI A. User profile model: a user dimension based classification[C]//Proceedings of the 10th international conference on intelligent systems: theories and applications. Rabat: IEEE, 2015.
[30] SCHÄFER C, ZINKE R, KÜNZER L, et al. Applying persona method for describing users of escape routes[J]. Transportation research procedia, 2014(2): 636－641.
[31] KAMTHAN P. Using personas to support the goals in user stories[C]// Proceedings of the 12th international conference on information technology-new generations. Las Vegas: IEEE, 2015.
[32] 陈志明,胡震云. UGC 网站用户画像研究[J]. 计算机系统应用,2017,26(1):24－30.
[33] 刘海鸥,黄文娜,谢姝琳. 面向情境化推荐服务的图书馆用户画像研究[J]. 图书馆学研究,2018(20):62－68.

[34] 刘速. 浅议数字图书馆知识发现系统中的用户画像——以天津图书馆为例[J]. 图书馆理论与实践,2017(6):103-106.
[35] 刘应杰. 基于数据挖掘的远程教育用户画像构建研究[D]. 西安:西安电子科技大学,2018.
[36] 李映坤. 大数据背景下用户画像的统计方法实践研究[D]. 北京:首都经济贸易大学,2016.
[37] 吴加琪. 我国用户画像研究的知识网络与热点领域分析[J]. 现代情报,2018,38(8):130-135,143.
[38] COOPER A, REIMANN R, CRONIN D, et al. About face 3.0: the essentials of interaction design[M]. New Jersey: John Wiley & Sons Inc, 2007.
[39] NIELSEN L. Ten steps to personas[EB/OL]. [2020-05-01]. https://www.personas.dk/2004/11/28/10-steps-to-personas/.
[40] MADUREIRA A, CUNHA B, PEREIRA J P, et al. Using personas for supporting user modeling on scheduling systems[C]//Proceedings of the 14th international conference on hybrid intelligent systems. Kuwait: IEEE, 2014.
[41] 关梓骜. 基于大数据技术的用户画像系统的设计与研究[D]. 北京:北京邮电大学,2018.
[42] 陈添源. 高校移动图书馆用户画像构建实证[J]. 图书情报工作,2018,62(7):38-46.
[43] 赵刚,姚兴仁. 基于用户画像的异常行为检测模型[J]. 信息网络安全,2017(7):18-24.
[44] 程全. 基于用户画像的数字图书馆信息服务模式研究[J]. 图书馆学刊,2018,40(4):68-71.

作者贡献说明:

宋雪雁:确定选题,提出论文研究框架,确定用户画像标签体系,修改论文;
张梦笛:收集论文资料,撰写论文,进行数据分析和处理。

Research on User Profile of Jinjiang Original Literature Website

Song Xueyan　Zhang Mengdi
School of Management, Jilin University, Changchun 130022

Abstract: [Purpose/significance] In order to expand the application scenarios of user profile theory, optimize Website operation, improve user stickiness, create greatest value for Jinjiang literature Website, intensive understand culture consumption of users and offer reference for research on users of other original literature Website, the paper explores the characteristics of Jinjiang original literature Website users and builds Jinjiang original literature website user profile. [Method/process] Research object of this paper targeted Jinjiang original literature website users, established user profiles of Jinjiang original literature Website and offered explanations related to the user interest and behavior by using exploratory factor analysis, K-Means cluster analysis and Word Art visualized analysis tools. [Result/conclusion] This paper obtains the user profile label system of Jinjiang original literature Website, establishes four typical user profiles of Jinjiang original literature Website, refers to different typical user profiles and puts forward optimization strategies for Jinjiang original literature Website.

Keywords: user profile　original literature Website　Jinjiang literature Website　user behavior